A curated Domain centric shared Docker registry linked to the Galaxy toolshed

François Moreews, Olivier Sallou, Yvan Le Bras, Grosjean Marie, Cyril Monjeaud, Thomas A Darde, Olivier Collin, Christophe Blanchet

A curated Domain centric shared Docker registry linked to the Galaxy toolshed

François Moreews[1], Olivier Sallou[2], Yvan le Bras[2], Marie Grosjean[3], Cyril Monjeaud[2], Thomas Darde[4], Olivier Collin[2], Christophe Blanchet[3]

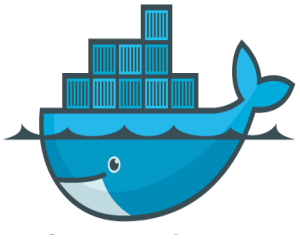
1 Genscale team -IRISA -Rennes, France

2 Genouest Bioinformatics facility – INRIA/IRISA – Rennes, France

3 French Institute of Bioinformatics – CNRS IFB-Core UMS3601 – Gif-sur-Yvette, France

4 INSERM U625 – Rennes France

Docker : presentation



“**Docker** is an open-source engine to easily create **lightweight, portable**, **self-sufficient containers** from any application. The same container that a developer **builds** and **test** on a laptop can run at **scale**, in **production**, on Vms,[...], public **clouds** and more.”

Docker : presentation

Why using Docker containers to build, deploy and execute applications ?

- **Efficient** (no virttualization)
- **Isolation**
- **Build one time, execute “anywhere”**, independently of the execution platform (laptops, clusters and clouds with linux kernels)

...

Build : Dependencies & Dockerfile

```
FROM ubuntu:12.0.4

ADD . /script

ENTRYPOINT ["perl /script/analysis.pl"]
```

Run Docker

```
docker -run
  containerUniqueID
 -k=31 -i input.fastq -o output.bam
```

-The docker run command acts as a wrapper of the tool command line.

-Host directories (input, output,work...) can be mounted inside the container.

Docker on the commercial Cloud

Google Container Engine

- A container based cloud architecture
- “With container-based computing, application developers can focus on their application code, instead of on deployments and integration into hosting environments”.

Docker on academic HPC clusters

- **Google Kubernetes** : an open source technology for containers life cycle management.
- **Docker Swarm** : allows to create and access to a pool of Docker hosts.
- **Genouest GO-DOCKER** : a batch scheduler like SGE, submitting jobs in Docker containers on top of Swarm..

Bioinformatics tools benchmarks with Docker

- cami-challenge.org : Critical Assessment of **Metagenomic** Interpretation
- http://nucleotid.es : continuous, objective and reproducible evaluation of genome **assemblers** using docker containers
- bioboxes.org : **interchangable** bioinformatics software containers

Galaxy Docker integration

- Docker can be used in Galaxy to :
  - manage tools dependencies : one tool , on Docker
  - Distribute populated Galaxy Distribution related to one topic

Shared registries : Docker Hub

- Not structured
- Not curated
- Not domain centric
- Not community driven

Shared registries : BIOSHADOCK

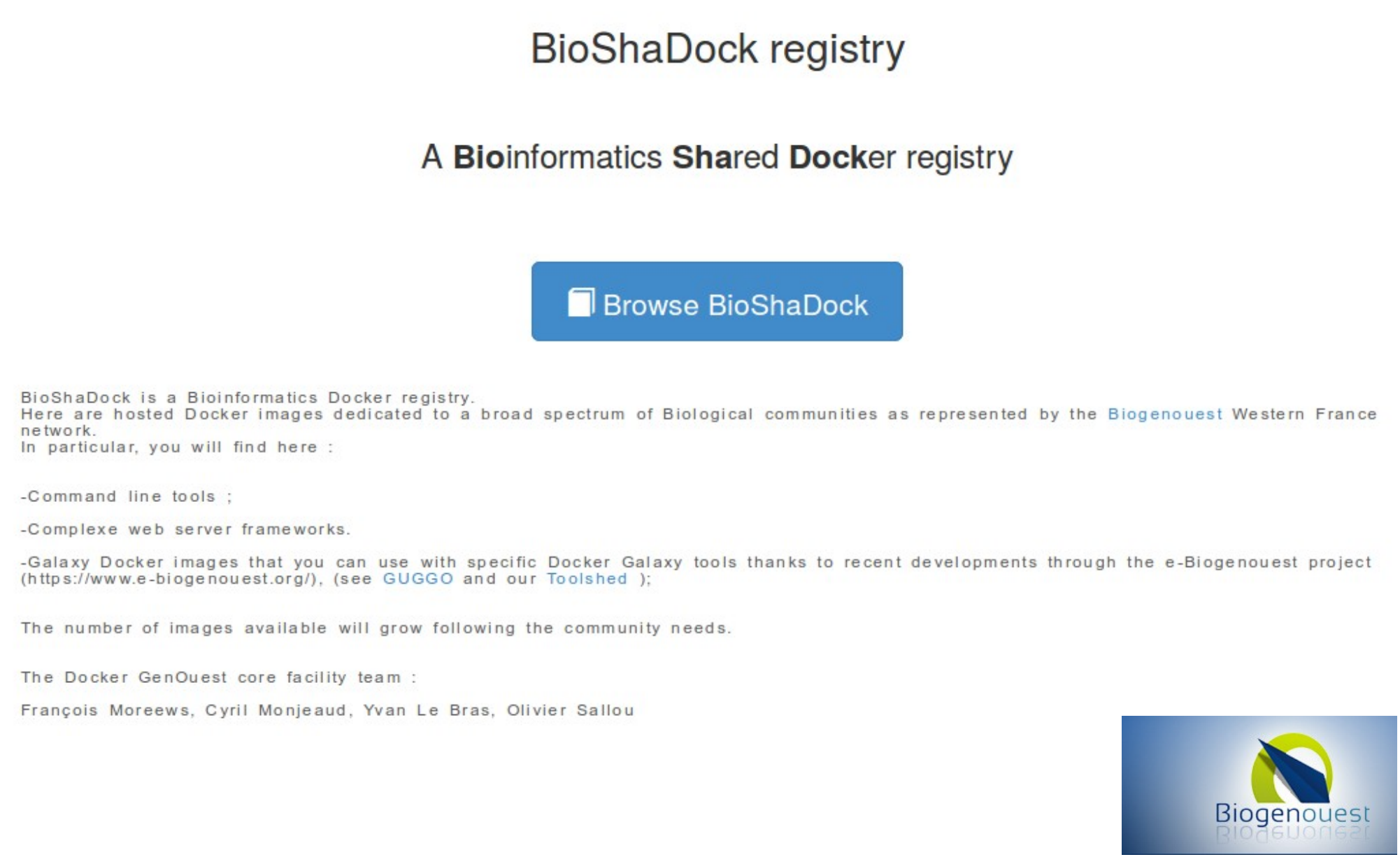
BIOSHADOCK

An initiative of the French Bioinformatics Institut & the Genouest Bioinformatics Facility

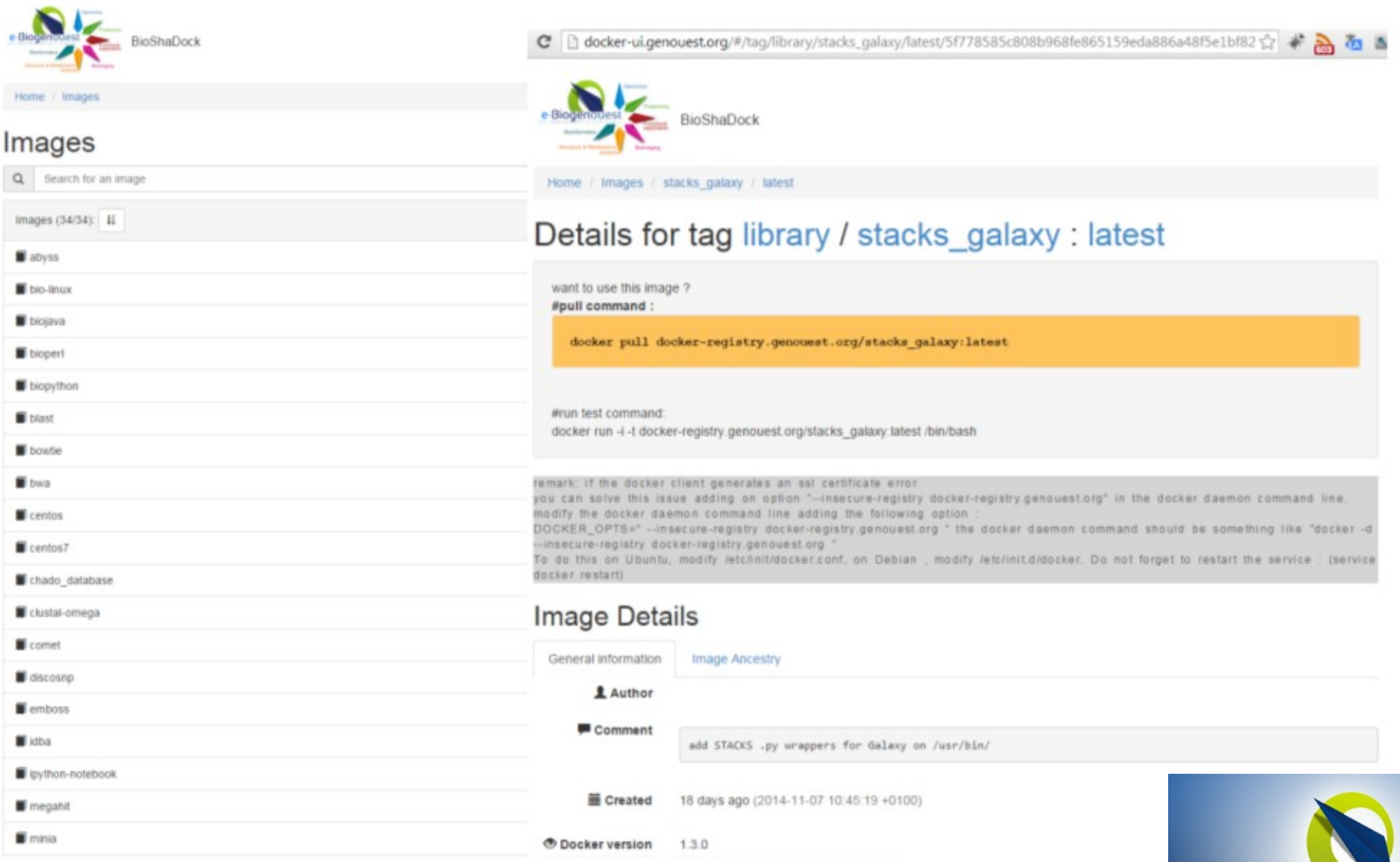
Goals :

- Federate bioinformatics tools deployment procedures for the IFB cloud infrastructure
- Generate customized Galaxy cloud instances on the fly.
- Docker image indexation (service registry & searches)

Shared registries : BIOSHADOCK

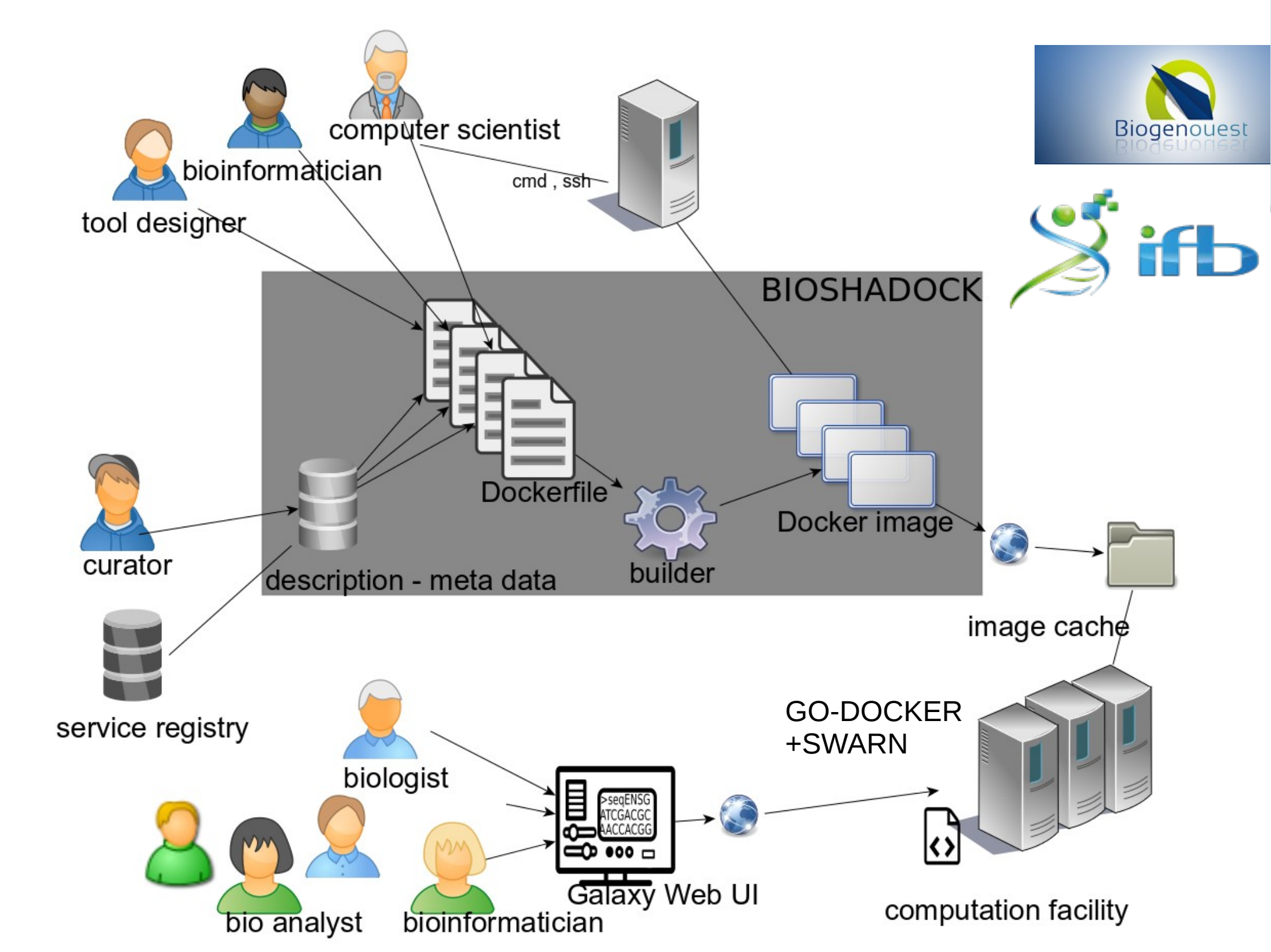


Shared registries : BIOSHADOCK

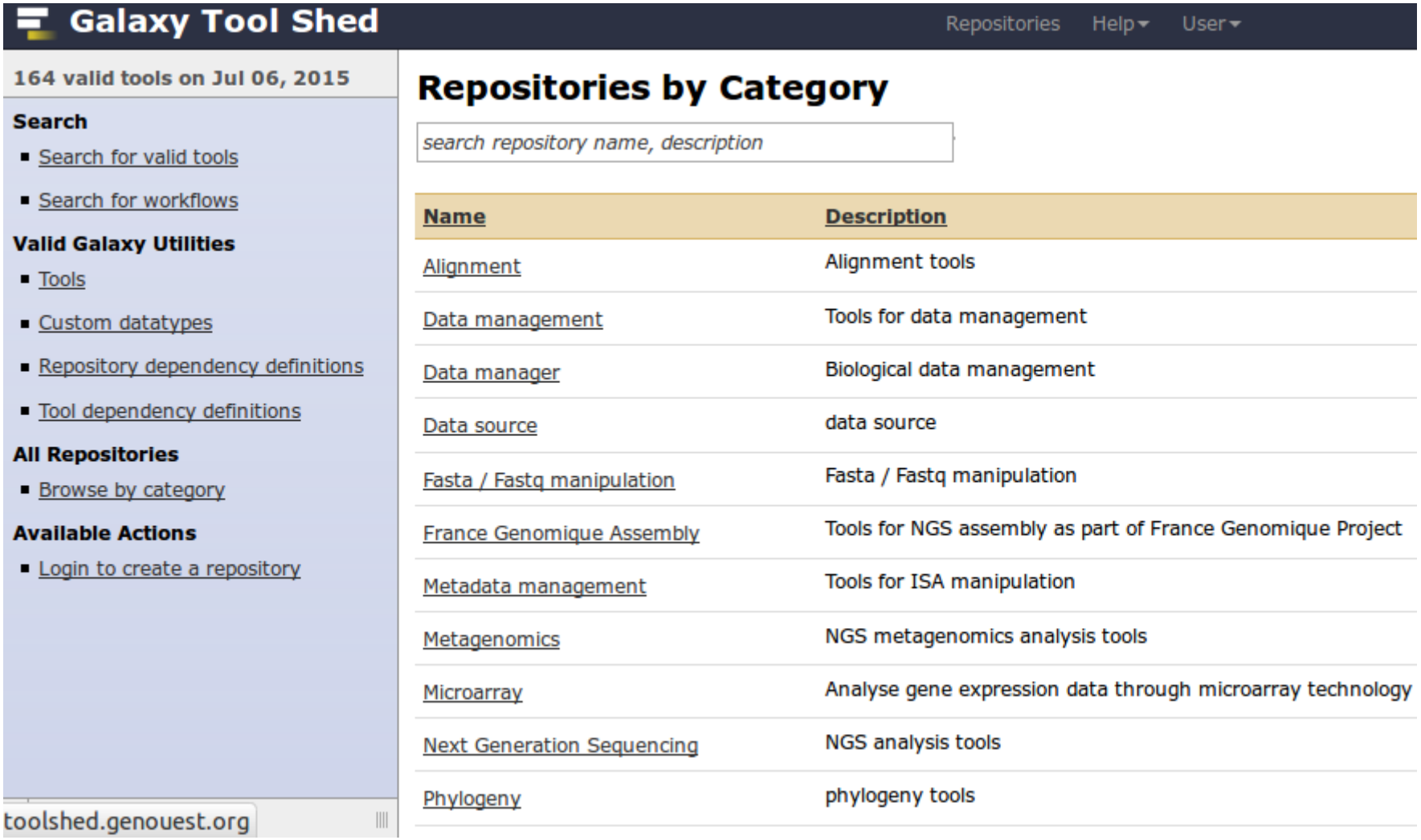


Shared registries : BIOSHADOCK

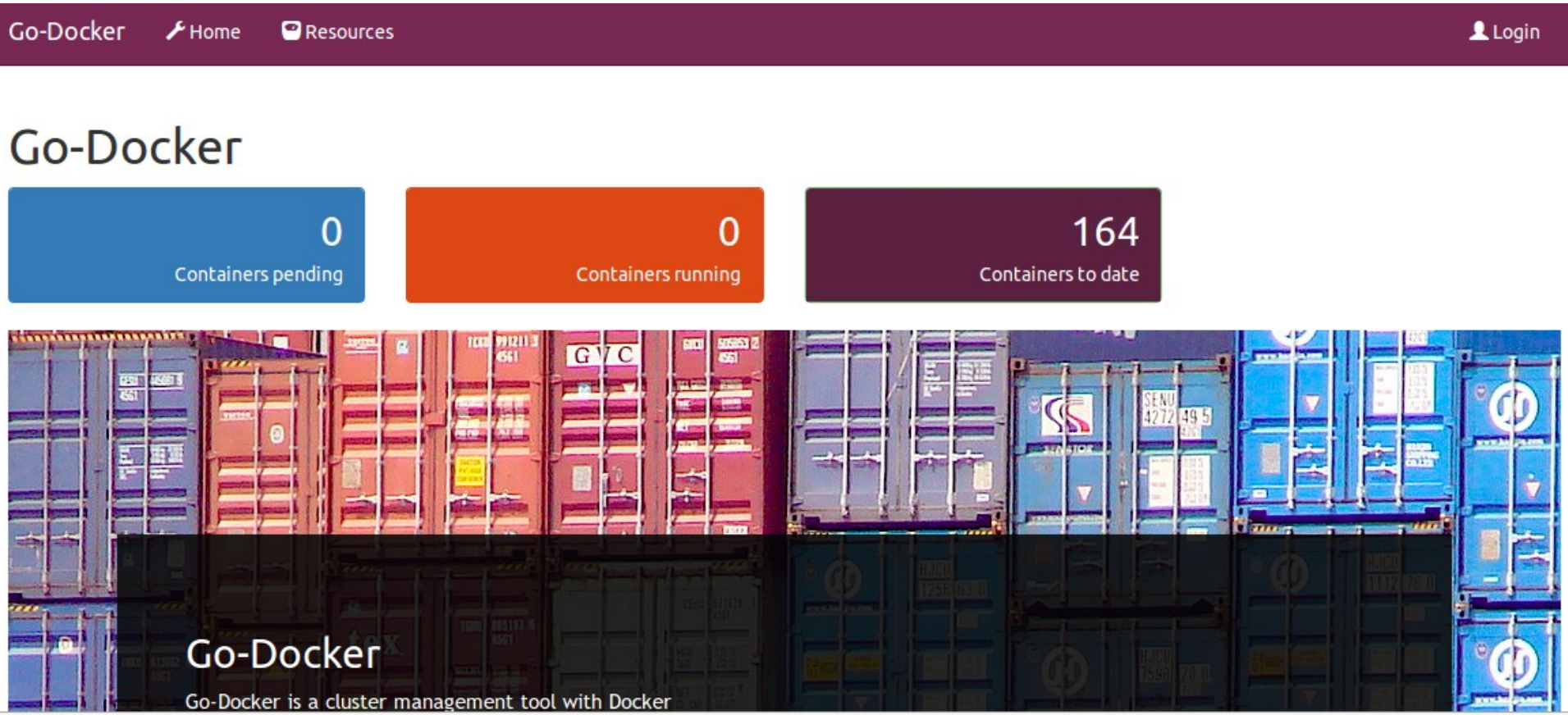
- BIOSHADOCK
  - Focuses on the model “on tool, one docker image”
  - Allows Dockerfile build
  - Manages permissions (private/ public images)
  - May integrate meta data to facilitate query and service registry searches
  - One unique repositity for softwares with or without tool.xml => SAAS + CMD
  - Integrated to Galaxy by redefining tools dependencies in a Toolshed


computer scientist
bioinformatician
tool designer
cmd , ssh
BIOSHADOCK
Dockerfile
Docker image
curator
description - meta data
builder
image cache
service registry
biologist
GO-DOCKER
+SWARN
>seqENSG
ATCGACGC
AACCACGG
Galaxy Web UI
bio analyst
bioinformatician
computation facility

BIOSHADOCK TOOLSHED integration

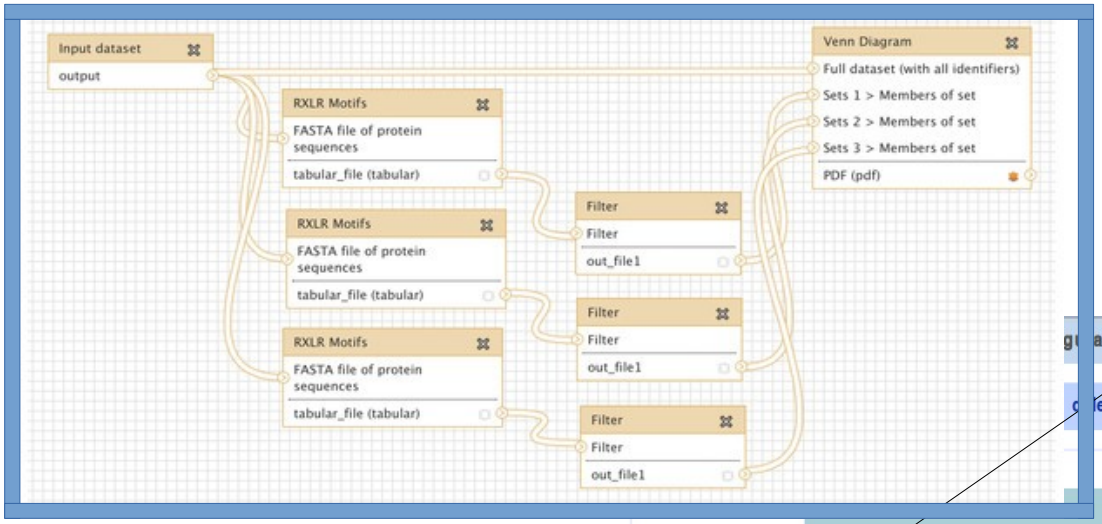


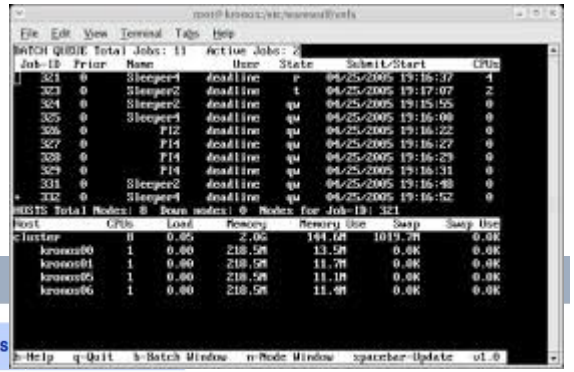
BIOSHADOCK cluster /cloud integration using GO-DOCKER



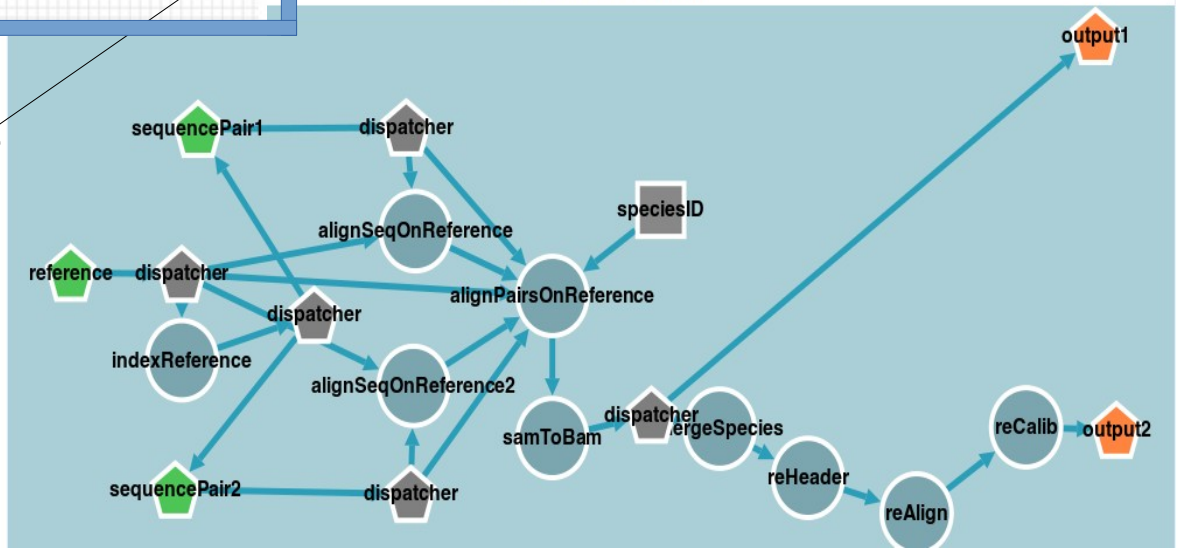
Build it one time, use it as you want

Command lines





Galaxy tools & workflows



BIOSHADOCK

Other SAAS tools

References

- Genouest GO-DOCKER :http://www.genouest.org/?p=246
- Google Kubernetes, Docker container cluster management : kubernetes.io
- BioShaDock, a Bioinformatics Shared Docker registry : http://docker-ui.genouest.org
- GUGGO Galaxy Tooshed : http://toolshed.genouest.org
- Nucleotid.es, continuous, objective and reproducible evaluation of genome assemblers using docker containers : http://nucleotid.es
- ELIXIR Tools and Data Services Registry : https://elixir-registry.cbs.dtu.dk
- Bioboxes, a standard for creating interchangable bioinformatics software containers : http://bioboxes.org
- IFB academic Cloud : http://www.france-bioinformatique.fr/?q=en/core/e-infrastructure-team/ifb-cloud